EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Simone Manganelli

Statistical decision functions with judgment

No 2512 / January 2021

Abstract

A decision maker tests whether the gradient of the loss function evaluated at a judgmental decision is zero. If the test does not reject, the action is the judgmental decision. If the test rejects, the action sets the gradient equal to the boundary of the rejection region. This statistical decision rule is admissible and conditions on the sample realization. The confidence level reflects the decision maker's aversion to statistical uncertainty. The decision rule is applied to a problem of asset allocation.

Keywords: Statistical Decision Theory; Hypothesis Testing; Confidence Intervals; Conditional Inference.

JEL Codes: C1; C11; C12; C13; D81.

Non-technical Summary

The use of judgment is ubiquitous in decision making, yet it lacks a formal treatment in statistics. Policy institutions, like central banks, routinely use state of the art econometric models to forecast key economic variables. When forecasts differ from the assessment of the decision makers, they are adjusted with 'expert judgment'.

The process of incorporating judgment in the decision process should be turned on its head. Decision makers should first express their judgmental decision and then econometricians should recommend whether there is statistical evidence to deviate from it. The statistical decision incorporating judgment lies at the boundary of a confidence interval.

Optimality is tested by checking whether, for a strictly convex loss function and confidence level, the gradient evaluated at the judgmental decision is statistically zero. Rejection of the null hypothesis implies that marginal moves decrease the loss function. This holds until the action associated with the closest boundary of the confidence interval of the gradient is reached. Abandoning a judgmental decision for a statistical procedure carries the risk of choosing a worse decision. The confidence level puts an upper bound to the probability of wrongly rejecting a decision when it is optimal.

The confidence level reflects the attitude of the decision maker towards statistical uncertainty. Decision makers who dislike any statistical uncertainty always follow their own judgmental decision and ignore the advice of the econometrician. At the other extreme, decision makers indifferent to statistical uncertainty ignore their judgment and always choose the maximum likelihood decision. Policy makers engaging in statistical decision making are likely characterized by low, but not extreme, confidence levels.

1 Introduction

The use of judgment is ubiquitous in decision making, yet it lacks a formal treatment in statistics. Policy institutions, like central banks, routinely use state of the art econometric models to forecast key economic variables. When forecasts differ from the assessment of the decision makers, they are adjusted with 'expert judgment', either by tinkering with the econometric model or by modifying directly the forecast. It is an *ad hoc* procedure, which clashes with the rigourous foundations of statistical decision theory.

The fundamental insight of this paper is that the incorporation of judgment in the decision process should be turned on its head. Decision makers should first express their judgmental decision and then econometricians should recommend whether there is statistical evidence to deviate from it. The logical conclusion is that the statistical decision incorporating judgment lies at the boundary of a confidence interval.

Optimality is checked by testing whether, for a strictly convex loss function and confidence level, the gradient evaluated at the judgmental decision is statistically equal to zero. Rejection of the null hypothesis implies that marginal moves decrease the loss function. This holds until the action associated with the closest boundary of the confidence interval of the gradient is reached. Abandoning a judgmental decision for a statistical procedure carries the risk of choosing a worse decision. The confidence level puts an upper bound to the probability of wrongly rejecting a decision when it is optimal.

Maximum likelihood decisions are obtained as a special case of this theory. Such decisions ignore judgment altogether, by setting the confidence level equal to one. In this case, the confidence interval degenerates into a single point, coinciding with the maximum likelihood estimate. The statistical interpretation is that the probability that the maximum likelihood decision produces a higher loss than the judgmental decision cannot be bounded away from one.

The contribution of this paper lies at the intersection between statistics and decision theory. Statistical decision theory emerged as a discipline in the 1950's with the works of Wald (1950) and Savage (1954). Berger (1985) provides a comprehensive and accessible review. Recent works within this tradition are Granger and Machina (2006), Patton and Timmermann (2012) and Elliott and Timmermann (2016). Other contributions include Chamberlain (2000) and Geweke and Whiteman (2006), who deal with forecasting using Bayesian statistical decision theory, and Manski (2013 and the references therein), who uses statistical decision theory in the presence of ambiguity for partial identification of treatment response. Manganelli (2009) introduces the original idea of incorporating judgment at the beginning of the decision process and proposes the heuristic decision rule of moving to the boundary of the confidence interval. The present paper provides a formal theoretical justification of such a procedure. It also shows that the decision incorporating judgment is admissible and therefore satisties the basic rationality principle of not being dominated by any other decision rule.

The paper is structured as follows. Section 2 sets up the decision environment and introduces the concept of judgment. Judgment is defined as a pair formed by a judgmental decision and a confidence level. It is used to define the hypothesis testing framework. The key results of this section are that the decision incorporating judgment is admissible, forms an essentially complete class and is either the judgmental decision itself or at the boundary of the confidence interval of the sample gradient of the loss function. The fundamental concept behind these results is *conditioning*. Section 3 provides economic intuition. The confidence level, by putting an upper bound to the probability of wrongly rejecting an optimal decision, can be interpreted as the preference of the decision maker toward statistical uncertainty. Section 4 uses an asset allocation problem as an illustrative example. Section 5 concludes.

2 Decisions with Judgment

This section introduces the concept of judgment and shows how hypothesis testing can be used to arrive at optimal decisions. I denote random variables with upper case letters (X) and their realization with lower case letters (x). The decision environment is formally defined as follows.

Definition 2.1 (**Decision Environment**).

1. *The observed data* $x^n \equiv (x_1', \ldots, x_n')'$ *are a realization from a p.d.f.* $f_t(x_t|\theta)$, *where* $x_t \in \mathbb{R}^v$, $\theta \in \mathbb{R}^p$, $v, p \in \mathbb{N}$. θ *is unknown.*
2. $\hat{\theta}(X^n)$ *is an extremum estimator (Newey and McFadden, 1994). The sample size* n *is sufficiently large, so that* $\sqrt{n}(\hat{\theta}(X^n) - \theta) \sim N(0, \Sigma)$.
3. $a \in \mathbb{R}^q$, $q \in \mathbb{N}$, *denotes the action of the decision maker.*
4. *The decision maker minimizes the loss function* $L(\theta, a)$, *which is strictly convex in* a *and twice continuously differentiable in* θ *and* a.

In words, Definition 2.1 assumes that 1) the econometric model is correctly specified, 2) the sample size is sufficiently large for asymptotic approximations to hold, 3) the action to be taken is any generic q-vector, and 4) a necessary and sufficient condition for an action to be optimal is to set the gradient of the loss function to zero.

2.1 Judgment

I introduce the following definition of judgment.

Definition 2.2 (**Judgment**)**.** ***Judgment*** *is the pair* $A \equiv \{\tilde{a}, \alpha\}$*.* $\tilde{a} \in \mathbb{R}^q$ *is the* ***judgmental decision****.* $\alpha \in [0,1]$ *is the* ***confidence level****.*

Judgment is routinely used in hypothesis testing, for instance when testing whether a regression coefficient is statistically different from zero (with zero in this case playing the role of the judgmental decision), for a given confidence level (usually 1%, 5% or 10%). I say nothing about how the judgmental decision is formed. This question is explored by Tversky and Kahneman (1974) and subsequent research. The choice of the confidence level is discussed in section 3. For the purpose of this paper, judgment is a primitive to the decision problem, like the loss function.

2.2 Hypothesis Testing

Define the action associated with the sample realization of the extremum estimator:

$$\hat{a}(x^n) \equiv \arg\min_a L(\hat{\theta}(x^n), a) \tag{1}$$

Define also the shrinkage action:

$$a(\lambda) \equiv \lambda \hat{a}(x^n) + (1-\lambda)\tilde{a} \qquad \lambda \in [0,1] \tag{2}$$

Notice that $\hat{a}(x^n)$, and therefore $a(\lambda)$, is observed at the time the decision is taken, as it depends on the sample realization.

The decision maker can test whether $a(\lambda)$ is optimal by testing if the gradient $\nabla_\lambda L(\theta, a(\lambda))$ is equal to zero. Define the gradient $Z(\theta, \lambda) \equiv \nabla_\lambda L(\theta, a(\lambda))$ to simplify notation. A test statistic for the gradient can be obtained by replacing θ with its extremum estimator $\hat{\theta}(X^n)$ and exploiting its asymptotic properties. The hypothesis to be tested is whether one should move in the direction of $a(1)$. Since by construction $Z(\hat{\theta}(x^n), 0) < 0$, $Z(\hat{\theta}(x^n), 1) = 0$ and by assumption 4 of Definition 2.1 is monotonic in λ, one can conclude that values of λ higher than 0 decrease the empirical loss function. The decision maker is interested, however, in the population value of the loss function. If the population gradient is positive, higher values of λ would increase the loss function, rather than decrease it. The null hypothesis to be tested is therefore that the population gradient has opposite sign relative to the sample gradient:

$$H_0 : Z(\theta, \lambda) \geq 0 \quad \text{vs} \quad H_1 : Z(\theta, \lambda) < 0 \tag{3}$$

The following theorem provides the distribution of the associated test statistic.

Theorem 2.1. ***(Test statistic)*** *Consider the decision environment of Definition 2.1 and define $Z(\theta, \lambda) \equiv \nabla_\lambda L(\theta, a(\lambda))$, where $a(\lambda)$ is defined in* (2)*. The distribution of the test statistic $Z(\hat{\theta}(X^n), \lambda)$ is*

$$\sqrt{n}\sigma^{-1}[Z(\hat{\theta}(X^n), \lambda) - Z(\theta, \lambda)] \sim N(0, 1) \tag{4}$$

where $\sigma^2 \equiv \nabla'_\theta Z(\theta, \lambda) \Sigma \nabla_\theta Z(\theta, \lambda)$.

Proof — See Appendix.

The random gradient $Z(\hat{\theta}(X^n), \lambda)$ depends not only on the random variable X^n, but also on the sample realization x^n via λ. Let us make this dependence explicit, by writing $Z(\hat{\theta}(X^n), \lambda(x^n))$. Under the null hypothesis $H_0 : Z(\theta, \lambda(x^n)) = 0$, the *p-value* is $\alpha_\lambda \equiv P(Z(\hat{\theta}(X^n), \lambda(x^n)) \leq Z(\hat{\theta}(x^n), \lambda(x^n)))$. The interpretation is the following. Repeating the hypothetical experiment of drawing independent values $\{x_h^n\}_{h=1}^H$ from the population distribution, the argument of the probability would be true α_λ of the times:

$$\lim_{H \to \infty} H^{-1} \sum_{h=1}^{H} I(Z(\hat{\theta}(x_h^n), \lambda(x^n)) \leq Z(\hat{\theta}(x^n), \lambda(x^n))) = \alpha_\lambda \tag{5}$$

When performing the thought experiment, $\lambda(x^n)$ is held fixed and does not change with x_h^n. By conditioning on the data, the potential realizations of the random variable X^n are not conflated with the observed realization x^n.

2.3 Decision

Consider the following standard definition (Wald, 1950):

Definition 2.3 (Decision Rule). *$\psi(X^n) : \mathbb{R}^{vn} \to \mathbb{R}^q$ is a **decision rule**, such that if $X^n = x^n$ is the sample realization, $\psi(x^n)$ is the action taken.*

In an hypothesis testing decision problem, only two actions are possible: the null hypothesis is either accepted or rejected. Recall that the hypothesis being tested is (3), that is whether the gradient evaluated at any given λ is non negative. Let $0 < \gamma < 1$ and $\Phi(c_\alpha) = \alpha$, where Φ is the cdf of the standard normal distribution. Define the decision rule $\psi^A(x^n, \lambda)$, where the dependence on the parameter λ has been made explicit:

$$
^{A}(x^n, \lambda) = \begin{cases} 0 & \text{if } \sqrt{n}\hat{\sigma}^{-1} Z(\hat{\theta}(x^n), \lambda) > c_{\alpha/2} \\ \gamma & \text{if } \sqrt{n}\hat{\sigma}^{-1} Z(\hat{\theta}(x^n), \lambda) = c_{\alpha/2} \\ 1 & \text{if } \sqrt{n}\hat{\sigma}^{-1} Z(\hat{\theta}(x^n), \lambda) < c_{\alpha/2} \end{cases} \tag{6}
$$

where $\hat{\sigma}^2 = \nabla'_\theta Z(\hat{\theta}(x^n), 1)\hat{\Sigma}\nabla_\theta Z(\hat{\theta}(x^n), 1)$ is a suitable estimator of the asymptotic variance. I have replaced $\nabla_\theta Z(\hat{\theta}(x^n), \lambda)$ with $\nabla_\theta Z(\hat{\theta}(x^n), 1)$ to eliminate the dependence of the variance on λ. This is possible, because under the null hypothesis that $a(\lambda)$ is optimal, $\hat{\sigma}^2$ is a consistent estimate of σ^2. Notice that since the sample gradient is negative by construction, the critical region is defined by $\alpha/2$, instead of α.

The test function (6) rejects the null hypothesis when it is equal to 1 and does not reject if it is equal to 0. The next theorem shows that this decision cannot be improved. Let us first report some additional definitions for convenience.

Definition 2.4 (**Risk Function**)**.** *The* ***risk function*** *associated with the decision rule $\psi(X^n)$ is $R(\theta, \psi) \equiv E_\theta(L(\theta, \psi(X^n)))$.*

Definition 2.5 (**Dominance**)**.** *A decision rule ψ_1* ***dominates*** *a decision rule $_2$ if $R(\theta, \psi_1) \leq R(\theta, \psi_2)$ for all $\theta \in \mathbb{R}^p$, with strict inequality for some θ. A rule ψ_1 is* ***equivalent*** *to ψ_2 if $R(\theta, \psi_1) = R(\theta, \psi_2)$ for all θ.*

Definition 2.6 (**Admissibility**)**.** *A decision rule is* ***admissible*** *if it is not dominated by any other decision rule.*

Definition 2.7 (**Completeness**)**.** *A class $\mathbb{C}$ of decision rules is* ***essentially complete*** *if, for any decision rule $\psi \notin \mathbb{C}$, there is a decision rule $\psi' \in \mathbb{C}$ which dominates or is equivalent to ψ.*

A standard result is that if an admissible decision rule ψ is not in an essentially complete class $\mathbb{C}$, then there exists a decision rule $\psi' \in \mathbb{C}$ which is equivalent to ψ (see for instance Lemma 2, page 522, Berger 1985). An essentially complete class does not necessarily contain all admissible decisions, but it contains all admissible risk functions.

Theorem 2.2. ***(Complete Class)*** *Given the decision environment of Definition 2.1, the decision rule associated with the test function $\psi^A(X^n, \lambda)$ in (6) is admissible and forms an essentially complete class.*

Proof — See Appendix.

The admissibility result is obtained by applying Karlin-Rubin theorem to the test function (6). It follows from two facts. First, even though the parameter θ may be a vector, the tested hypothesis is about a scalar. Second, the randomness of the decision rule stems from the test function $\psi^A(X^n, \lambda)$, while the corresponding actions in case of rejection or non rejection are not random.

The next theorem finally derives the decision compatible with judgment:

Theorem 2.3. ***(Decision with judgment)*** *Consider the decision environment of Definition 2.1. A decision maker with judgment $A = \{\tilde{a}, \alpha\}$ selects the action $a(\hat{\lambda})$ from (2), where $\hat{\lambda}$ is the max between 0 and the solution that sets the test function $\psi^A(X^n, \hat{\lambda}) = \gamma$ in (6).*

Proof — See Appendix.

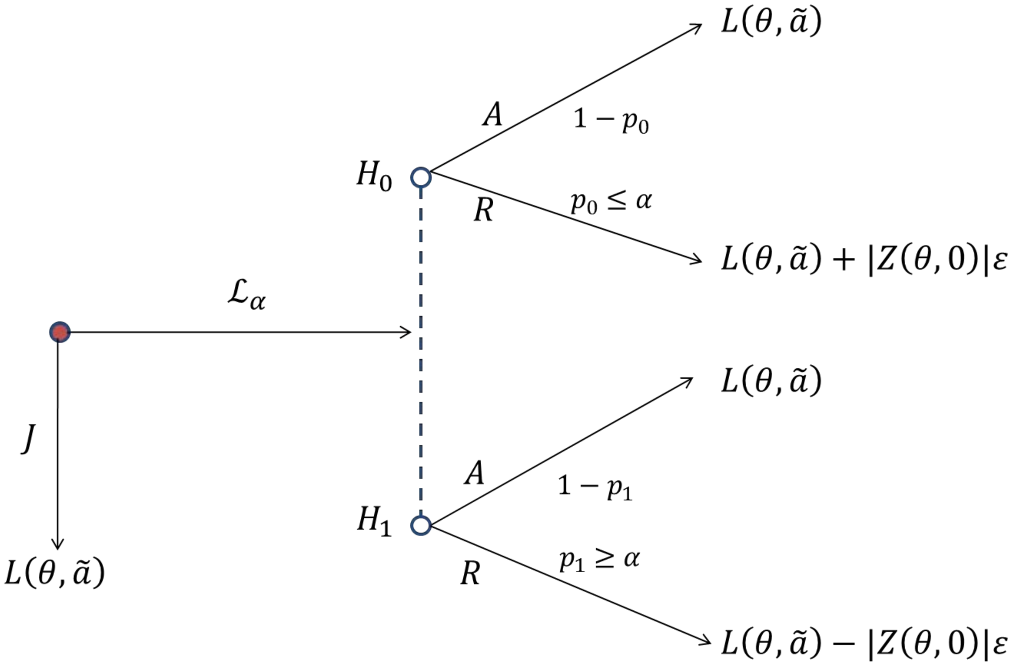
3 Economic Intuition

To understand the intuition behind Theorem 2.3, consider that the null hypothesis (3) for $\lambda = 0$ is a statement about the population gradient evaluated at the judgmental decision $\tilde{a}$. It says that *marginal* moves from $\tilde{a}$ in the direction of $\hat{a}(x^n)$ do not decrease the loss function. If it is not rejected at the given confidence level α, the chosen action must be $\tilde{a} = a(0)$. Rejection of the null hypothesis, on the other hand, implies accepting the alternative, which states that *marginal* moves away from $\tilde{a}$ decrease the loss function.

The decision problem is depicted in Figure 1. The decision maker has two possible choices: 1) hold on to the judgmental decision $\tilde{a}$, denoted by the action J, or 2) follow the econometrician's advice, which is equivalent to accepting the bet $\mathcal{L}_\alpha$. In this second case, there is no information to distinguish the upper part of the decision tree, denoted by the node H_0, from the lower part, denoted by H_1. Under H_0, the null hypothesis (3) is true, so that any deviation from the judgmental decision $\tilde{a}$ does not result in a lower loss. A marginal $\varepsilon > 0$ move away from $\tilde{a}$ results in the loss $L(\theta, \tilde{a}) + |Z(\theta, 0)|\varepsilon$ for sufficiently small ε. Under H_1, the null hypothesis is false, so that a marginal ε move away from $\tilde{a}$ results in the loss $L(\theta, \tilde{a}) - |Z(\theta, 0)|\varepsilon$. The dash line connecting the two nodes represents true uncertainty for the decision maker, in the sense that it is not possible to attach any probability to being in H_0 or in H_1. The decision maker can choose the confidence level α, which puts an upper bound to the probability that the null is wrongly rejected when it is true, and a lower bound to the probability of correctly rejecting H_0 when it is false.

In case of rejection, the preferred decision $a(\hat{\lambda})$ is the action which lies at the boundary of the $(1-\alpha)$-confidence interval of the gradient $Z(\hat{\theta}(X^n), \hat{\lambda})$. Other actions would not be compatible with the confidence level α of the decision maker. In fact, actions closer to the original judgmental decision $\tilde{a}$ are rejected at the confidence level α, while actions further away may be wrongly rejected with a probability greater than α.

Figure 1: Statistical Decision Tree



Note: A decision maker with judgmental decision $\tilde{a}$, confidence level α and loss function $L(\theta,\tilde{a})$ can choose $\tilde{a}$ (branch J) or follow a statistical decision rule (branch $\mathcal{L}_\alpha$). For a given estimate $\hat{\theta}(x^n)$ of the statistical parameter, the rule tests whether marginal ($\varepsilon > 0$) deviations from $\tilde{a}$ are warranted. It will not decrease the loss if $\tilde{a}$ is optimal (node H_0) and it will not increase the loss if $\tilde{a}$ is not optimal (node H_1). The dashed line connecting H_0 and H_1 represents uncertainty, as the decision maker cannot distinguish between the two parts of the tree and no probability can be attached to them. By choosing α, she can control the probability p_0 of increasing the loss function, in case H_0 is true. α provides also the lower bound to the probability p_1 of correctly deviating from $\tilde{a}$ in case H_1 is true.

The confidence level α determines the willingness of the decision maker to engage in the statistical bet. A decision maker who likes statistical uncertainty chooses $\alpha = 100\%$. When $\alpha = 100\%$ the confidence interval degenerates into a single point and the null hypothesis that $\tilde{a}$ is optimal is always rejected. This corresponds to the *maximum likelihood* decision $a(1)$. At the other end of the spectrum, a decision maker with an extreme aversion to statistical uncertainty chooses $\alpha = 0\%$. When $\alpha = 0\%$ the confidence interval degenerates into the entire real line and the null hypothesis that $\tilde{a}$ is optimal is never rejected. This corresponds to the *minmax* decision $a(0)$. An intermediate case is represented by the *subjective classical* estimator of Manganelli (2009), which sets $\alpha \in (0, 1)$ and gives the decision $a(\hat{\lambda})$ of Theorem 2.3.

There is a trade-off associated with the choice of the confidence level, which is the one associated with Type I and Type II errors in hypothesis testing. Lower values of α imply a lower probability of wrongly rejecting the null hypothesis, but also a lower probability of correctly rejecting it. It is up to the decision maker to solve this trade-off. The choice of the confidence level depends on the decision problem at hand and the confidence that decision makers have on their own judgmental decision. Notice that it is impossible not to choose. Any decision maker facing a statistical decision problem is forced to choose a confidence level.

4 An Asset Allocation Example

This section implements the decision with judgment, solving a standard portfolio allocation problem.[1] The empirical implementation of the mean-variance

[1] See Gelain and Manganelli (2020) for an application to estimation of Dynamic Stochastic General Equilibrium models.

asset allocation model introduced by Markowitz (1952) has puzzled economists for a long time. Despite its theoretical success, standard estimators of the portfolio weights produce volatile asset allocations with poor out-of-sample performance (Brandt 2007). This paper takes a different perspective on this problem, by starting with a judgmental decision and testing whether its performance can be improved.

To implement the statistical decision rule of Theorem 2.3, I take a monthly series of closing prices for the EuroStoxx50 index, from February 1987 until September 2019. EuroStoxx50 covers the 50 leading Blue-chip stocks for the Eurozone. The data is taken from DataStream. The closing prices are converted into period log returns, $x^n \equiv (x_1, \ldots, x_n)'$, for a total of $n = 392$ monthly observations. Assume for simplicity $E_t(X_{t+1}) = \theta_1$ and $V_t(X_{t+1}) = \theta_2$, that is both first and second moments are not time varying, and define $\theta \equiv [\theta_1, \theta_2]'$. The methodology can be readily applied to cases where the conditional mean and variance are time varying.

Consider an investor with a quadratic utility function $U(W) = W - 0.5bW^2$ with $b > 0$ and $W < 1/b$. The decision is about the fraction a of cash W_n to be invested in the stock market. Assuming that the return on cash is zero, the monthly portfolio returns are ax_i, for $i = 1, \ldots, n$. The loss function is the negative of the expected utility and is, up to a linear transformation, $L(\theta, a) = -(1 - bW_n)a\theta_1 + 0.5bW_n a^2(\theta_2 + \theta_1^2)$. The decision associated with the maximum likelihood estimate is $\hat{a}(x^n) = (1 - bW_n)\hat{\theta}_1/(bW_n(\hat{\theta}_2 + \theta_1^2))$, where $\hat{\theta}_1 = n^{-1} \sum_{i=1}^n x_i$ and $\hat{\theta}_2 = n^{-1} \sum_{i=1}^n (x_i - \hat{\theta}_1)^2$. The robust asymptotic variancecovariance matrix is $\hat{\Sigma} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$ (see Newey and McFadden, 1994), where:

$$\hat{A} = n^{-1}\sum_{i=1}^{n} \nabla_\theta s(x_i, \hat{\theta}), \qquad \hat{B} = n^{-1}\sum_{i=1}^{n} s(x_i, \hat{\theta}) s(x_i, \hat{\theta})'$$

$$s(x_i, \hat{\theta}) = [(x_i - \hat{\theta}_1)\hat{\theta}_2^{-1}, -0.5\hat{\theta}_2^{-1} + 0.5(x_i - \hat{\theta}_1)^2\hat{\theta}_2^{-2}]'$$

$$\nabla_\theta s(x_i, \hat{\theta}) = \begin{bmatrix} -\hat{\theta}_2^{-1}, & -(x_i - \hat{\theta}_1)\hat{\theta}_2^{-2} \\ -(x_i - \hat{\theta}_1)\hat{\theta}_2^{-2}, & 0.5\hat{\theta}_2^{-2} - (x_i - \hat{\theta}_1)^2\hat{\theta}_2^{-3} \end{bmatrix}$$

The gradient is $Z(\hat{\theta}, \lambda) = -(1 - bW_n)(\hat{a}(x^n) - \tilde{a})\hat{\theta}_1 + bW_n a(\lambda)(\hat{a}(x^n) - \tilde{a})(\hat{\theta}_2 + \hat{\theta}_1^2)$ and $\nabla_\theta Z(\hat{\theta}, 1) = [-(1 - bW_n)(\hat{a}(x^n) - \tilde{a}) + 2bW_n a(1)(\hat{a}(x^n) - \tilde{a})\hat{\theta}_1, bW_n a(1)(\hat{a}(x^n) - \tilde{a})]'$. There are now all the elements to compute $\hat{\sigma}^2$.

The decision rule of Theorem 2.3 is implemented by choosing $bW_n = 0.9$ and the choices of $A = \{\tilde{a}, \alpha\}$ reported in Table 1. By choosing $\alpha = 100\%$, the decision maker ignores any judgmental decision and selects the decision associated with the maximum likelihood estimate. In the current exercise, this corresponds to investing 19% of the portfolio in the stock index and keeping the rest in cash. At the other extreme, by choosing $\alpha = 0\%$, the decision maker ignores any statistical evidence and selects the judgmental decision. This can be seen from the fact that the decisions in the column under $\alpha = 0\%$ are identical to the corresponding $\tilde{a}$.

Intermediate choices of the confidence level, $\alpha \in (0, 1)$, result in decisions which shrink toward the maximum likelihood decision, provided there is sufficient statistical evidence to move away from the judgmental decision. The null hypothesis that $\tilde{a} = 0$ is optimal is not rejected at the 10% confidence level, and therefore the decision coincides with $\tilde{a}$. Notice that this finding explains the lack of participation in the stock market, even though the standard expected utility theory predicts that all agents should always invest some fraction

Table 1: Asset allocation decisions

		α			
		0%	1%	10%	100%
	0	0	0	0	0.19
$\tilde{a}$	0.5	0.5	0.5	0.4	0.19
	1	1	0.53	0.4	0.19

Note: Share of wealth invested in the monthly Eurostoxx50 index, according to alternative choices of judgmental decision ($\tilde{a}$) and confidence level (α). $\alpha = 100\%$ always ignores any judgmental decision and chooses the decision associated with the maximum likelihood estimate. $\alpha = 0\%$ always ignores any statistical evidence and chooses the judgmental decision. $\alpha \in (0, 1)$ results in decisions which shrink toward the maximum likelihood decision, provided there is sufficient statistical evidence to move away from the judgmental decision.

of their wealth in the risky asset. Investors averse to statistical uncertainty prefer not to invest in the stock market if the available statistical evidence is not strong enough. This explanation is consistent with a larger body of literature which explains the lack of participation with the assumption that investors view stock returns as ambiguous (Epstein and Schneider, 2010).

Investors prefer also not to move away from the judgmental decision $\tilde{a} = 0.5$ at the 1% confidence level. This judgmental decision is however rejected at 10%, leading to an investment in the stock market of 40% of the overall portfolio. Notice that $\tilde{a} = 1$ is also rejected and leads to the same decision as the one associated with $\tilde{a} = 0.5$.

5 Conclusion

Judgment plays an important role not just for individuals, but also in policy institutions. Most policy decisions are shaped by the judgment of policy makers. When advising a policy maker, the econometrician can test whether the preferred judgmental decision is supported by models and data. If not, the decision incorporating judgment is always at the closest boundary of the confidence interval. The probability of wrongly rejecting the judgmental decision is bounded by the given confidence level. The decision rule is admissible and is obtained by properly conditioning on the observed sample realization.

The confidence level reflects the attitude of the decision maker towards statistical uncertainty. Decision makers who dislike any statistical uncertainty always follow their own judgmental decision and ignore the advice of the econometrician. At the other extreme, decision makers indifferent to statistical uncertainty ignore their judgment and always choose the maximum likelihood decision. Policy makers engaging in statistical decision making are likely characterized by low, but not extreme, confidence levels.

Appendix — Proofs

Proof of Theorem 2.1 — Consider a mean value expansion of the test statistic:

$$Z(\hat{\theta}(X^n), \lambda) = Z(\theta, \lambda) + \nabla'_\theta Z(\bar{\theta}(X^n), \lambda)(\hat{\theta}(X^n) - \theta)$$

The result follows from assumption 2 of Definition 2.1. $\square$

Proof of Theorem 2.2 — Define $\mathcal{X} \equiv \sqrt{n}\hat{\sigma}^{-1}Z(\hat{\theta}(X^n), \lambda)$ and $\vartheta \equiv \sqrt{n}\sigma^{-1}$ $Z(\theta, \lambda)$. Given assumption 2 in the decision environment of Definition 2.1 that n is finite, but sufficiently large for the asymptotic approximation to hold, it follows that $\mathcal{X} \sim N(\vartheta, 1)$, under the null hypothesis that $a(\lambda)$ is optimal. The optimality test for $a(\lambda)$ can be equivalently rewritten as $H_0 : \vartheta \geq 0$.

The normal distribution belongs to the exponential family and therefore possesses a monotone likelihood ratio (see section 1 of Karlin and Rubin, 1956). The test function foresees two actions: the null hypothesis is either accepted or rejected. In case of rejection, it prescribes to marginally move from $a(\lambda)$ towards $a(1)$. Denote the marginal move with $a(\lambda + \varepsilon)$, for $\varepsilon > 0$. Since the loss function is unique up to a positive linear transformation, let $L_1 \equiv \sqrt{n}\sigma^{-1}\varepsilon^{-1}L(\theta, a(\lambda))$ and $L_2 \equiv \sqrt{n}\sigma^{-1}\varepsilon^{-1}L(\theta, a(\lambda + \varepsilon))$ be the losses corresponding to the two actions. Taking the limit for $\varepsilon \to 0$:

$$\begin{aligned}\lim_{\varepsilon \to 0}(L_2 - L_1) &= \sqrt{n}\sigma^{-1}Z(\theta, \lambda) \\ &\equiv \vartheta\end{aligned}$$

The function $\lim_{\varepsilon \to 0}(L_2 - L_1)$ changes sign only once as a function of ϑ. The conditions of Theorem 5, page 530, of Berger (1985) are satisfied and the result follows. □

Proof of Theorem 2.3 — If $\psi^A(x^n, 0) = 0$, the null hypothesis $H_0 : Z(\theta, 0) \geq 0$ is not rejected at the given confidence level α. $\tilde{a}$ is therefore retained as the chosen action.

If $\psi^A(x^n, 0) = 1$, the null hypothesis is rejected. Let $\hat{\Delta}$ be the value satisfying $\sqrt{n}\hat{\sigma}^{-1}Z(\hat{\theta}(x^n), \hat{\Delta}) = c_{\alpha/2}$. Given the convexity assumption 4 in Definition 2.1, this value exists and is unique. Denote with $a(\hat{\lambda})$ the chosen action and suppose by contradiction that $\hat{\lambda} \neq \hat{\Delta}$. If $\hat{\lambda} < \hat{\Delta}$, this implies that $\sqrt{n}\hat{\sigma}^{-1}Z(\hat{\theta}(y^n), \hat{\lambda}) < c_{\alpha/2}$, that is $H_0 : Z(\theta, \hat{\lambda}) \geq 0$ is rejected. This decision cannot be optimal. If $\hat{\lambda} > \hat{\Delta}$, continuity implies that it exists $\varepsilon > 0$ such that the null $H_0 : Z(\theta, \hat{\lambda} - \varepsilon) \geq 0$ is rejected at the given confidence level α, even though $\sqrt{n}\hat{\sigma}^{-1}Z(\hat{\theta}(y^n), \hat{\lambda} - \varepsilon) > c_{\alpha/2}$. This decision cannot be optimal. The chosen action must therefore be $\hat{\lambda} = \hat{\Delta}$. □

References

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer-Verlag.

Brandt, M.W. (2009), Portfolio Choice Problems, in Y. Ait-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*, North Holland.

Chamberlain, G., (2000), Econometrics and decision theory, *Journal of Econometrics*, 95 (2), 255-283.

G. Elliott and A. Timmermann (2016), *Economic Forecasting*, Princeton University Press.

Epstein, L.G. and M. Scnheider (2010), Ambiguity and asset markets, *Annual Review of Financial and Economics*, 2, 315–346.

Gelain, P. and S. Manganelli (2020), Monetary policy with judgment, ECB Working Paper No. 2404.

Geweke, J. and C. Whiteman (2006), Bayesian Forecasting, in G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol.1, 81-98, Elsevier.

Granger, C.W.J. and M.J. Machina (2006), Forecasting and decision theory, in G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol.1, 81-98, Elsevier.

Karlin, S. and H. Rubin (1956), The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio, *The Annals of Mathematical Statistics*, 27, 272-299.

Manganelli, S. (2009), Forecasting with Judgment, *Journal of Business and Economic Statistics*, 27 (4), 553-563.

Manski, C.F. (2013), *Public policy in an uncertain world: analysis and decisions*, Harvard University Press.

Markowitz, H.M. (1952), Portfolio Selection, *Journal of Finance*, 39, 47-61.

Patton, A.J. and A. Timmermann (2012), Forecast Rationality Tests Based on Multi-Horizon Bounds, *Journal of Business and Economic Statistics*, 30 (1), 1-17.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, John Wiley & Sons.

Tversky, A. and D. Kahneman (1974), Judgment under Uncertainty: Heuristics and Biases, *Science*, 1124-1131.

Wald, A. (1950), *Statistical Decision Functions*, New York, John Wiley & Sons.

Acknowledgements

The views expressed in this paper are my own and do not necessarily reflect those of the ECB.

Simone Manganelli

European Central Bank, Frankfurt am Main, Germany; email: simone.manganelli@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-4512-7 ISSN 1725-2806 doi:10.2866/29005 QB-AR-21-003-EN-N